DOCUMENT RESUME

ED 393 859 TM 024 462

AUTHOR	Bond, Linda Ann; Roeber, Edward D.
TITLE	The Status of State Student Assessment Programs in the United States. Annual Report.
INSTITUTION	Council of Chief State School Officers, Washington, D.C.; North Central Regional Educational Lab., Oak Brook, IL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	NCREL-RPIC-SSAP-AR-95
PUB DATE	Jun 95
CONTRACT	RP91002007
NOTE	40p.; For related documents, see TM 024 463-464.
AVAILABLE FROM	North Central Regional Educational Laboratory, Order Department, 1900 Spring Road, Suite 300, Oak Brook, IL 60181 ($9.95, order number RPIC-SSAP-AR-95; diskettes for Windows on MacIntosh also available).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cooperation; *Databases; *Educational Assessment; Educational History; Educational Trends; Elementary Secondary Education; Performance Based Assessment; State Programs; *Student Evaluation; Surveys; *Testing Programs; *Test Use; Trend Analysis
IDENTIFIERS	Alternative Assessment; Large Scale Assessment; *Large Scale Programs; *State Student Assessment Program Database

ABSTRACT

In 1991 the collaborative efforts of the Council of Chief State School Officers and the North Central Regional Educational Laboratory resulted in the current form of the State Student Assessment Program (SSAP) database. This report marks the third year of that partnership, which builds on earlier data collection efforts to present information about large-scale assessment programs at the state level. The survey collects information annually to describe state programs in traditional testing and nontraditional assessment and the uses made of these assessments. Chapter 2 describes state assessment programs. Chapter 3 explores newer forms of assessment, and Chapter 4 reviews additional assessment issues. Chapter 5 considers the history and trends in statewide assessment. Appendixes present definitions of terms, a summary table of state assessment programs, and the order form for the products of the SSAP. (Contains 16 charts, 2 tables, and 2 figures.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 859

The Status of State Student Assessment Programs in the United States

Annual Report, June 1995

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Council of Chief State School Officers

NCREL
North Central Regional Educational Laboratory

BEST COPY AVAILABLE

TM024482

©1995 North Central Regional Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

RPIC-SSAP-AR-95 $9.95

THE STATUS OF STATE STUDENT ASSESSMENT PROGRAMS IN THE UNITED STATES

THE ANNUAL REPORT OF THE STATE STUDENT ASSESSMENT PROGRAMS DATABASE

June 1995

Linda Ann Bond, Ph.D., and Edward D. Roeber, Ph.D.
with assistance from Diane King and David C. Braskamp

Council of Chief State School Officers

North Central Regional Educational Laboratory

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS
One Massachusetts Ave NW, Suite 700
Washington, D.C. 20001-1431

Gordon Ambach, Executive Director
Ramsay Selden, Director, State Education Assessment Center

The Council of Chief State School Officers (CCSSO) is a non profit organization of the 57 public officials who head departments of public education in every state, U.S. territory, and the District of Columbia. CCSSO seeks its members' consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public. Because the Council represents the chief education administrator in each state and territory, it has access to the educational and governmental establishments in each state, and the national influence that accompanies this distinct position. CCSSO forms coalitions with many other educational organizations, including those organizations that are active in assisting the nation and the states in setting high standards for their students and those that assess the performance of students against these high standards.

The State Education Assessment Center provides a central clearinghouse to improve data acquisition, monitoring, and the assessment of education. More recently, the State Collaborative on Assessment and Student Standards (SCASS) was formed to network states and other groups to develop prototype and complete assessment components for a variety of content areas. Projects are taking place in a number of areas. The goal in all of these projects is to encourage the development of higher quality student assessments at lower cost to the states. The Council also supports the Association of State Assessment Programs (ASAP), an informal network of the assessment staffs in the states.

NCREL
North Central Regional Educational Laboratory

NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY
1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480

Jeri Nowakowski, Executive Director
Deanna Durrett, Director, Regional Policy Information Center

The North Central Regional Educational Laboratory (NCREL) helps education professionals in a seven-state region support school restructuring to promote learning for all students, especially those most at risk.

One of ten federally supported educational laboratories, NCREL responds to the needs of educators in Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin in the critical program areas of curriculum, instruction, and assessment; early childhood and family education; professional development; rural education; and urban education.

NCREL's Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance, technology policy, and student assessment policy. RPIC publishes *Policy Briefs* on a variety of topics including charter schools, decentralization, interagency collaboration. Policy Seminars are conducted annually in cooperation with each state served.

NCREL also houses the Midwest Regional Center for Drug-Free Schools and Communities, one of five federally funded centers that provide training, dissemination, special products, and other activities to prevent alcohol, tobacco, and other drug use among youth.

BEST COPY AVAILABLE

Acknowledgments

Special thanks go to all the State Assessment Directors who make the State Student Assessment Programs Database possible by providing rich information about their assessment programs. Thanks also to the Chief State School Officers who supported us in this effort from its inception.

We could not have managed this project without the tireless efforts of Deb Roeber, who spent numerous hours on the phone "nudging" those who were a little late returning the survey, and Dina Czocher, who followed through to see that all was progressing smoothly.

Last but not least, thank you to Arie van der Ploeg, Diane King, and David Braskamp of NCREL, who spent hours ensuring the accuracy of the data, creating data tables in Paradox, designing and building the paper and electronic books within which the data tables are housed, and creating the charts and graphs for this report. They were also valuable reviewers of drafts, and discovered creative ways to present and analyze the data, all of which improved the quality of the report considerably. As always, this was a team effort.

Table of Contents

Chapter One: Introduction to the State Student Assessment Program Database 1

Chapter Two: Overview of State Student Assessment Programs 3
- Number of States with an Assessment Program 3
- Number of Assessment Components Per State 3
- Types of Assessments Used by States 3
- Purposes for Statewide Assessments 4
- Assessment Consequences 6
- Subject Areas Assessed 7
- Grade Levels Assessed 8
- Summary 9

Chapter Three: Newer Forms of State Assessment 10
- Types of Nontraditional or Alternative Items 11
- Constraints on Developing Nontraditional Assessment Items 13
- Writing Assessment 14
- Summary 15

Chapter Four: Additional Assessment Issues 16
- Sampling 16
- Calculator Use 17
- Assessment of Special Populations 18
- Developmental Process for State Tests 19

Chapter Five: History and Trends in Statewide Assessment 21
- Criterion-Referenced Assessment and Minimal Competency Tests 21
- Advent of Writing Assessment 22
- Expansion to Other Subject Areas 22
- Performance Assessment 22
- Professional Development on Assessment 23
- Norm-Referenced Tests 23
- National Efforts at Joint Development 24
- Future Issues and Their Impact on State Assessment 25

Appendices
- Definition of Terms 28
- Summary Table of State Assessment Programs 29
- State Student Assessment Program Database Order Form 30

List of Charts

Chart 2-1a: Number of Assessment Components 4
Chart 2-1b: Number of Students Tested, by Test Type 4
Chart 2-2: Patterns of Assessment in State Assessment Systems 5
Chart 2-3: Number of Assessment Purposes Per State 5
Chart 2-4: Major Assessment Purposes 6
Chart 2-5: Major Subjects Assessed by States 8
Chart 2-6: Grades States Assess 8

Chart 3-1: Major Subjects Assessed by States with Nontraditional Items 11
Chart 3-2a: Nontraditional Exercise Types Used by States: Language Arts and Writing 12
Chart 3-2b: Nontraditional Exercise Types Used by States: Mathematics and Science 12
Chart 3-3a: Developmental Stages of Nontraditional Items: Mathematics and Writing 12
Chart 3-3a: Developmental Stages of Nontraditional Items: Science and Language Arts 13

Chart 4-1: Types of Assessment Sampling 16
Chart 4-2: Changes in Calculator Use for Math and Science 17
Chart 4-3: Accommodations for Students With Special Needs 18
Chart 4-4: Services Provided by Contractors 19

List of Tables

Table 2-1: Number of Assessment Components 3
Table 2-2: States with a Graduation Test 7

List of Figures

Figure 2-1: Assessment Patterns Across the United States 4
Figure 2-2: States with High School Graduation Tests 7

Chapter One

Introduction to the State Student Assessment Programs Database

The topic of student assessment generates considerable controversy among educators and members of the public. Some view large-scale assessment programs as a critical element of the reform and change needed in American schools. Two primary reasons for this are: (1) assessment can provide direction and motivation to students, parents, teachers, and others to help students learn the skills needed to succeed both in school and in life after school; and (2) assessment programs can help gauge the success of our schools. An indication of the strength of their appeal is the number of states that have student assessment programs: 47. The three states that do not have testing programs are Iowa, Nebraska, and Wyoming.[1]

However, there are those educators and members of the public who do not view large-scale assessments positively. Critics feel such programs exert negative pressure on teachers and students. Much of the debate surrounds such issues as the content covered by the assessment, the type of assessment used, how the assessment is scored, and the uses made of the assessment results. Whether viewed positively or negatively, large-scale assessment programs are a fact of life in most states in the United States.

While state assessment programs share some common purposes and methods, they can also be quite different. Differences exist for various reasons—for example, the educational policy climate in the state, the technical quality issues surrounding the use of assessment to make high-stakes decisions, or the status of curricular reform in the state. We need to recognize these differences in order to understand the assessment programs that exist and the options that are available to change these programs.

In addition, we need to recognize the movement in Washington to limit the federal role in education. A result of this has been that states likely will have more control over the educational resources provided to their schools. Therefore, state assessment practices will continue to play a major role in educational reform.

The Association of State Assessment Programs (ASAP), an informal organization of state assessment directors, began collecting information about large-scale assessment programs at the state level in 1977. The results of the annual ASAP surveys were provided to states in the form of a written summary of each state's assessment program. In 1991 Ed Roeber, ASAP's chairperson, became director of student assessment programs for the Council of Chief State School Officers (CCSSO). A partnership with the North Central Regional Educational Laboratory (NCREL) led to the current form of the State Student Assessment Program (SSAP) database. This report is a result of the third year of that partnership.

[1] Colorado and Massachusetts suspended their assessment programs temporarily in 1993-1994. Nebraska is in the process of developing its first state assessment program.

As the information deepens with time, we are able to provide more meaningful information to states because we are able to monitor patterns of change in state assessment programs. As data collection continues in the future, we hope to sharpen the analysis of change in statewide assessment practices.

The survey annually collects three kinds of information. Part One of the survey asks each state to describe its existing program, its collaborative partners, and what it is developing. Part Two of the survey asks each state to describe its efforts in nontraditional assessment and, this year, in high school graduation testing. Part Three of the survey asks each state to divide its assessment program into components, or sets of assessments, that are used to gather data for different assessment purposes. For each component, states explain who is tested, what subjects are tested, and what types of assessments are used. From this detail, we can build an accurate picture of what statewide assessment programs look like and how they are attempting to accomplish their state assessment goals. This report is a summary to provide an understanding of what the 50 states are doing and how they are doing it.

Chapter Two

Overview of State Student Assessment Programs

This chapter provides an overview of the assessment the states conduct. A tabular overview appears in the Summary Table in the Appendix. The detailed responses for each state to the survey are available in the companion publication *State Student Assessment Programs Database, June 1995.*

Number of States With an Assessment Program

Statewide assessment programs are almost universal. In the 1993-1994 school year, 45 of the 50 states conducted some form of statewide assessment. Colorado and Massachusetts temporarily suspended their assessment programs while developing new ones. Nebraska is at work developing its first assessment program, to be implemented before 1998. Iowa and Wyoming continue to be the only states that report no state-mandated assessment program in place or in development.

Number of Assessment Components Per State

State assessment programs are typically multifaceted. We felt that it was critical that states define and describe each unique component in detail. In the survey, we defined a component as a single assessment or group of assessments that share a common purpose or set of purposes. Much of the information that is provided in subsequent sections of this report comes from these component descriptions. Table 2-1 lists the number of components for each state.

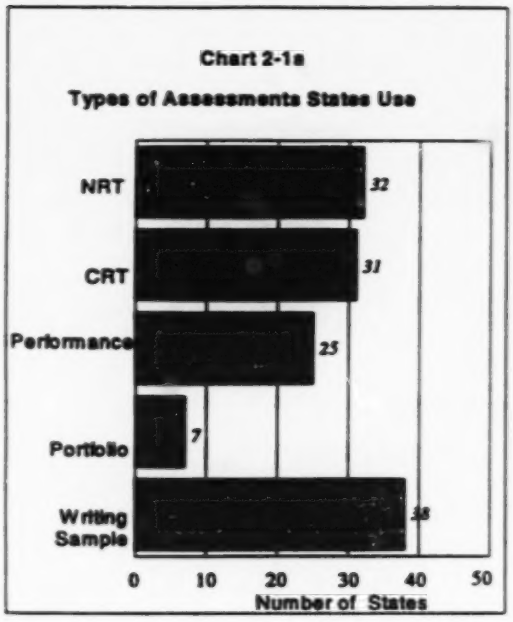
Types of Assessments Used by States

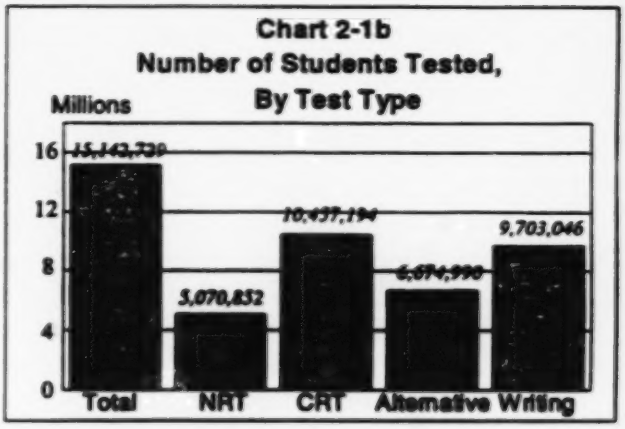
Chart 2-1a shows the number of states reporting the use of norm-referenced assessments, criterion-referenced assessments, writing assessments, performance events, and portfolios. Writing samples continue to be the most widespread form of assessment, used by 38 states. The number of states reporting criterion-referenced assessments decreased from 33 in 1992-1993 to 31 in 1993-1994, while the number of states using norm-referenced assessments increased by one to 32 in 1993-1994. The number of states with performance-based assessments continued to grow, from 17 in 1991-92, to 23 in 1992-93, to 25 in 1993-94. States using portfolios remained constant at seven.

Table 2-1

Number of Assessment Components

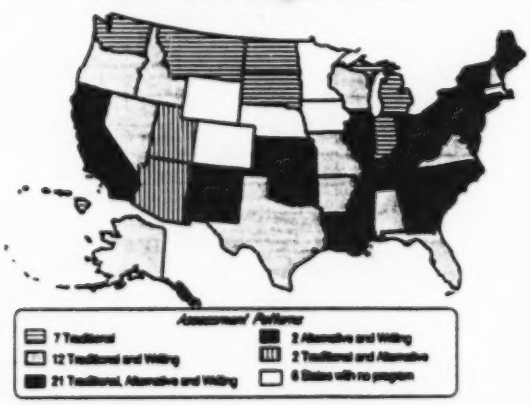
State		State		State		State		State	
AK	2	HI	3	ME	1	NJ	2	SD	2
AL	7	IA	0	MI	2	NM	4	TN	4
AR	2	ID	2	MN	1	NV	4	TX	1
AZ	2	IL	1	MO	2	NY	7	UT	3
CA	3	IN	1	MS	3	OH	4	VA	2
CO	0	KS	1	MT	1	OK	2	VT	2
CT	2	KY	3	NC	1	OR	2	WA	1
DE	2	LA	4	ND	1	PA	2	WI	2
FL	3	MA	0	NE	0	RI	3	WV	3
G	6	MD	3	NH	1	SC	2	WY	0





However, examining the number of students actually tested by each type of test presents a very different picture. (See Chart 2-1b.) About twice as many students take CRTs as take NRTs, even though NRTs are given in one more state than CRTs. The number of students providing writing samples is also quite large--just below that given CRTs. These numbers show a clear trend towards CRTs, writing, and alternative assessments, and away from NRTs.

Figure 2-1
Assessment Patterns Across the United States



Most states conduct several types of assessment programs. Figure 2-1 shows the pattern of assessment types across the states.

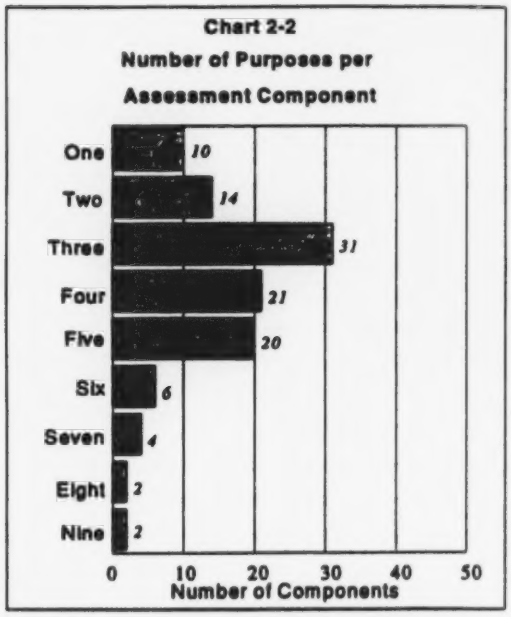
The most common pattern, evident in 17 states, includes three types of assessment: traditional, nontraditional or alternative, and writing samples. We define traditional assessments as consisting of multiple-choice tests, including norm-referenced tests (NRTs) and criterion-referenced tests (CRTs), while nontraditional assessments include performance tasks and/or portfolios.[2] Seven states have only traditional assessments while two states conduct only alternative assessments coupled with writing samples.

Purposes for Statewide Assessments

Most states use each of their assessment components for two to five purposes, as may be seen in Chart 2-2. This situation creates tensions for students, teachers, and schools, especially if some of the purposes are seen to conflict.

[2]An NRT yields comparisons against a normative group, while a CRT assesses performance against stated outcomes. Performance tasks and portfolios are intended to provide more in-depth coverage of important learner outcomes that cannot be well measured using traditional, multiple-choice tests.

BEST COPY AVAILABLE



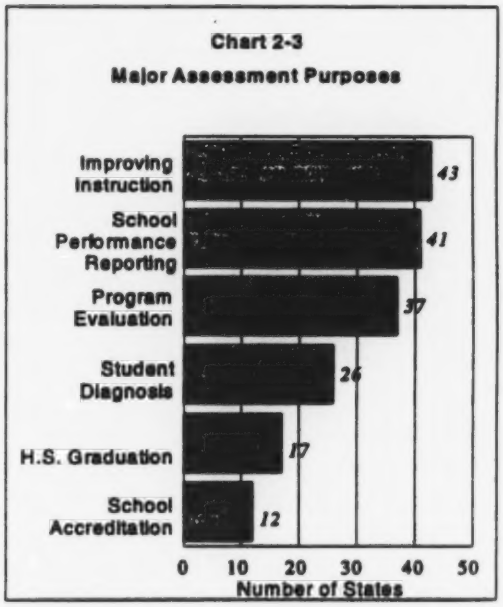


Chart 2-3 displays the six most common purposes states cite for assessing student performance. School and student purposes are much more common than teacher purposes. Only Tennessee reports using one of its assessment components for teacher evaluation (New York allows districts to do so if they choose). With respect to individual student purposes, 17 states use assessments for high school graduation tests (two less than in 1992-1993), and 26 for student diagnosis (one less than 1992-1993). The top three overall assessment purposes—improvement of instruction and curriculum, school performance monitoring (a form of accountability), and program evaluation—all are school or programs. Thirty-four states, approximately 75 percent of the states with assessment programs, each operate at least one assessment component that has all three of these purposes. Forty states, or 89 percent, have at least one component for which both accountability and instructional improvement are cited.

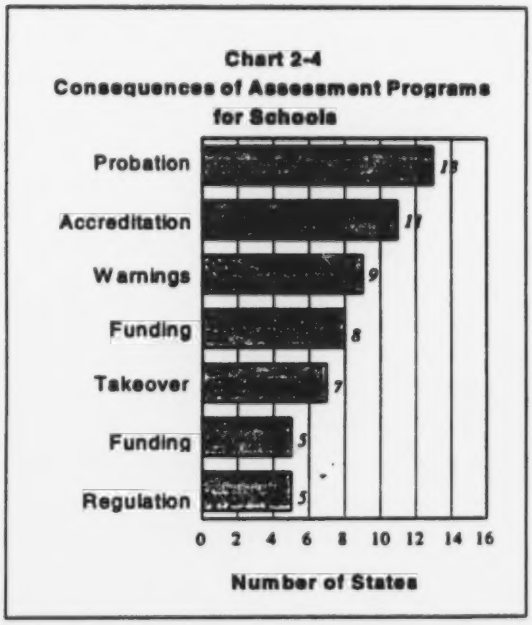
As discussed earlier, states depend on assessments to meet many purposes, but some combinations of purposes create more tension than others. Attempting to use a state assessment program for school or student accountability and for instructional improvement can be especially problematic. Designing an assessment program to meet high-stakes accountability purposes typically requires standardization of content, administration, and scoring. Accuracy of scoring and standardization of procedure is paramount, particularly if a high school diploma may be denied based on a student's score. Test security is high, with results determined at a centralized scoring center and returned weeks, sometimes months, after the assessment is administered.

The very safeguards that ensure comparability and fairness limit the utility of the results for instructional decision-

BEST COPY AVAILABLE

making. For an assessment to be effective as an instructional improvement tool, the results need to be made available almost immediately so teachers can adjust their instruction. Reviewing assessment results over the summer may be helpful for curriculum planning, but teachers need access to ongoing assessment information to modify instructional strategies within the classroom. A classroom-based assessment system, albeit somewhat standardized by virtue of the learning goals being assessed, requires continuous, unobtrusive collection of assessment data, flexible administration, and immediate feedback. Unfortunately, this flexibility, vital to classroom assessment, is typically seen to violate the standardization necessary for accountability purposes.

The state assessment directors acknowledge the difficulty inherent in using one assessment program for both accountability and instructional improvement purposes. However, law and regulation often require they do so. States, therefore, are designing assessment systems that try to capture both sets of purposes in ways to minimize the conflict between them. Some states, such as Illinois, are developing assessment systems with layers at the state and local levels that are aligned to the same learner goals, but used for different purposes. The state assessment serves accountability purposes primarily, while the local assessments are used for instructional improvement and school improvement planning. Other states, such as Vermont, are combining regionalized scoring of some student assessments with intensive teacher inservice to improve the accuracy of classroom portfolios for use as potential accountability data. Still others, Kentucky, for example, are auditing the results of local assessments to ensure that scoring guidelines are being applied uniformly across the state to improve comparability of scores.



The primary goal of state assessment continues to be the improvement of instruction in order to help students meet new, challenging standards. But, states seem unsure whether improved assessment content and format or increased accountability will result in the most improvement. They therefore continue to do both, a situation that limits the utility of the assessment program for either purpose.

Assessment Consequences

This year's survey asked also about the consequences of assessment results for schools, staff, and students. Chart 2-4 displays the most common consequences identified for schools. These can be quite severe. Some combination of funding gains and losses, loss of accreditation status, warnings, and eventual takeover of

BEST COPY AVAILABLE

schools are potential consequences in 23 states.

Currently, consequences for school staff are much less common, with two states reporting financial awards, one state reporting financial penalties, and one state reporting probation.

Consequences for students remain fairly rare also. Six states report basing student promotion decisions on state assessments, and ten states make student award and recognition decisions based on their assessments.

High school graduation tests, however, are another matter.[1] Figure 2-2 shows the 18 states that conducted high school graduation tests in 1993-1994.[2]

Table 2-2 categorizes the states by the requirements they place on students to graduate from high school, to receive an endorsement on their diploma, or to receive an honors diploma. These tests are the ones that most often end up in court. In order to successfully defend against a lawsuit, careful attention must be paid to the content of the test (it must match what has been taught), the timing of the notice (students need to know approximately three years ahead of time that passing the exam will be a requirement for graduation), and the technical quality of the exam (the test must be reliable, valid, and fair).[3]

Figure 2-2
States With High School Graduation Tests

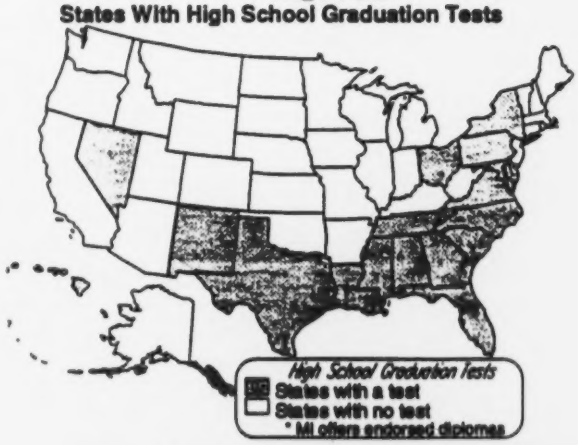


Table 2-2

States With a Graduation Test

Graduation			
Alabama	Louisiana	New	Ohio
Florida	Maryland	New	South Carolina
Georgia	Mississippi	Nevada	Tennessee
Hawaii	North	New York	Texas
			Virginia
Endorsed			
Michigan	New York	Tennessee	
Honors			
New York	Ohio	Tennessee	

Subject Areas Assessed

Five subjects are likely to be assessed by states no matter what assessment is used (see Chart 2-5).

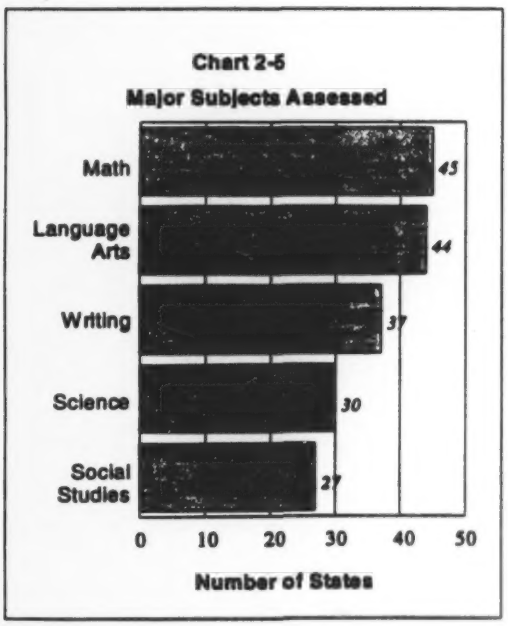
All the states with assessment programs assess mathematics; language arts (including reading) is assessed in every state but one. Writing is assessed in 36 states, down from 39 last year. There was also a slight drop in science (down from 34 states in 1991-92 to 30 states in 1993-94) and social studies (down from 29 states to 27). These decreases

[1] A complete report on these data will be released in the fall.

[2] This is two fewer than last year, due to recategorization of the data. This area continues to be volatile. For instance, Michigan's high school diploma endorsement is no longer considered a graduation test (students do not have to pass to graduate). Indiana is implementing a high school graduation test as we write.

[3] For a discussion of the legal issues involved with such tests, see: Phillips, S. (1993) ***Legal Implications of High Stakes Assessment***, Oak Brook, Illinois: North Central Regional Educational Laboratory.

BEST COPY AVAILABLE



may indicate situations where assessment cost is becoming a factor.

Other subjects, such as music, foreign languages, health, vocational education, visual arts, and physical education, are assessed by fewer than five states apiece.

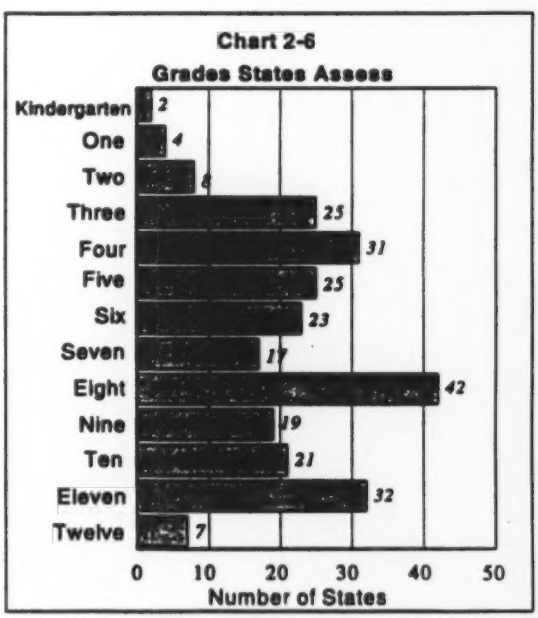
Subjects appear not to be assessed separately for purposes of accountability and improvement of instruction. Assessment in these five subjects most often follows the pattern of multiple purposes; in each subject area, almost all assessments are used for both accountability and instructional improvement.

Grade Levels Assessed

Which grades and how many grades are assessed varies widely among statewide assessment programs and components. Some patterns are worth mentioning, however. States are least likely to assess students in the early primary grades. States are most likely to assess students in grades 4, 8, and 11, as shown in Chart 2-6. All forms of assessment tend to be administered at these benchmark grades. Forty-two of the 45 states with assessment programs assess in the 8th grade, and 31 and 32 assess at the 4th and 11th grade levels respectively. A review by assessment types results in the following general findings:

- Norm-referenced assessments clearly peak at benchmark grades 4, 8, and 11.
- Criterion-referenced assessments also peak at these benchmark grades, but are also frequently given at the grade levels between.
- Performance assessments show a similar grade-level pattern as NRTs.
- Portfolios are given in too few states to detect a pattern.

Writing samples also occur most at the benchmark grades, but with a particularly strong peak at grade 8.



Summary

Over the past three years, certain findings of the survey have been consistent. State assessment remains a significant tool for educational reform in 45 states. In general, students are assessed most often at grades 4, 8, and 11 for the purposes of improvement of instruction, school accountability or school performance reporting, and program evaluation. At grades four and eight, roughly half of all students nationally are assessed at least once each year by their state. Approximately one-third of the states with assessment programs require students to pass an exam to graduate. Students are assessed most often with a combination of traditional and alternative assessments with few states relying on traditional or alternative assessments alone. The use of alternative assessments in conjunction with traditional assessments continues to grow.

The tensions that exist when assessment is used for both school or student accountability and instructional improvement continue to cause difficulty for those who design and implement these programs. Unfortunately, most states require these conflicting purposes in their programs. The tensions are often further complicated by placing negative consequences on poor performance, thus increasing the stakes for schools and students.

Chapter Three

Newer Forms of State Assessment

States continue to explore alternatives to the traditional multiple-choice assessments that have been and continue to be the most popular form of assessment in state assessment programs. About one-half of the states with assessment programs are using performance events to enhance traditional, multiple-choice assessments. Only two states report that they are using alternative assessments exclusively: Kentucky, which uses performance assessments and portfolios; and Maine, which uses performance assessments, portfolios, and writing assessments. Four other states rely heavily on nontraditional assessments. California relies primarily on alternative assessments, although some multiple-choice assessments also are used.

Arizona reports the use of a norm-referenced test alongside the state's major assessment program, which includes performance assessments, portfolios, and writing assessments. Vermont primarily uses mathematics and writing portfolios, but also administers uniform tests in mathematics (a short, criterion-referenced test) and a uniform assessment in writing (a writing sample). Maryland retains a traditional seventh-grade functional literacy test, but its major assessment program consists of performance assessments and writing samples. A similar number of states rely exclusively on traditional assessment: Hawaii, Indiana, Michigan, Montana, North Dakota, South Dakota, and Washington.

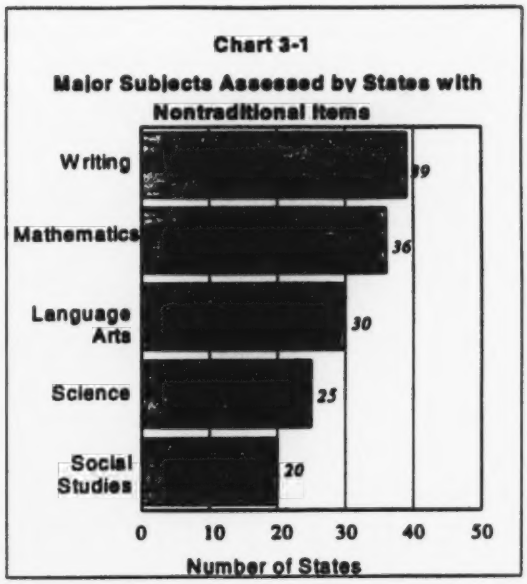
The use of alternative assessment in conjunction with traditional assessment is growing. This practice is in part due to changes in student standards—what students should know and be able to do. Changes in the workplace and in the skills needed for life in an information age suggest that students need knowledge and skills that will enable them to solve increasingly complex problems. Many of these skills cannot be assessed using traditional, multiple-choice assessment, and this is causing many states to explore alternatives. These alternatives usually become additions to traditional assessments.

Multiple-choice assessments require students to select a "right" answer from among several "wrong" answers. While this form of assessment is certainly useful for assessing knowledge and is often considered a direct application of knowledge, open-ended assessments that require students to generate their own solutions to assessment problems or tasks are becoming increasingly necessary to assess new learner outcomes. Many states are concerned that relying exclusively on traditional assessments results in a narrowed curriculum that produces students who memorize a lot of facts and skills, but have little ability to apply them to real-life situations. One of the major benefits of nontraditional assessment is that, in addition to judging the correctness of the student's answer, the

appropriateness of the procedure that the student employed is also considered.

Since no assessment type is ideal for all purposes and content, nontraditional assessments have their trade-offs—most notably, the increased cost and time associated with their development, administration, and scoring. Ensuring the reliability of the assessment results has also proven costly and difficult, although the benefits in improved assessment of complex skills and the modeling of good instruction is worthwhile to some states. For these reasons, and because traditional assessments still can measure some learner outcomes well, most states are not completely replacing their traditional assessment programs with nontraditional assessments (see Figure 2-1 in Chapter 2). Rather, they are adding nontraditional programs to traditional programs, which also are getting a face lift with new content and standards. Another difficulty of nontraditional assessments is generalizability. Different performance tasks evoke different levels of skill from different students. This limits the likelihood that a given performance on a small sample of tasks will be strongly indicative of the student's overall ability.

State activity in the development of nontraditional exercises in all subjects is depicted in Chart 3-1. Nontraditional assessment activity is up in all subjects, with the most development activity apparent in writing, mathematics, other language arts (including reading), science, and social studies. The number of projects within states is even more interesting. In most of these states, four to ten developmental projects are underway in three to six subject areas. The states are

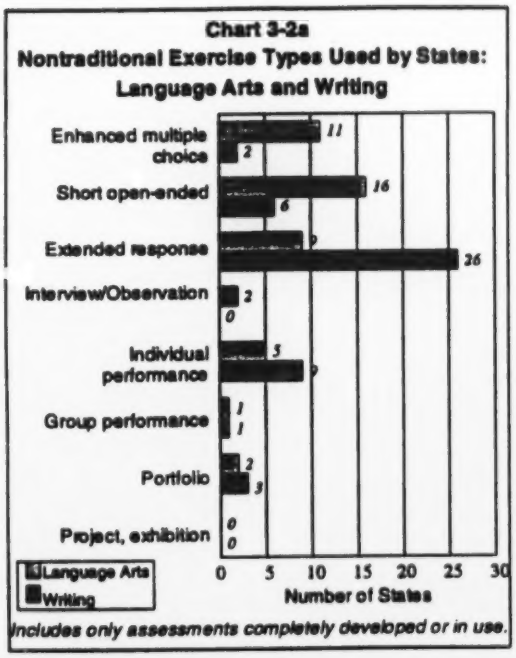


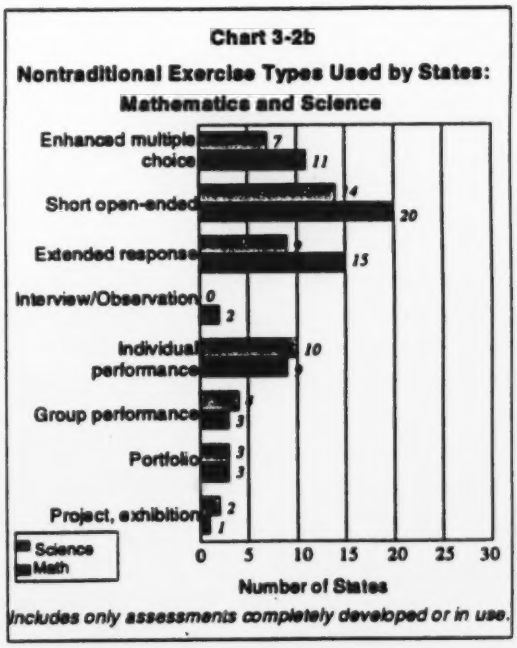
continuing to demonstrate clear interest in expanding their assessment options.

Types of Nontraditional or Alternative Items

The desire to improve the quality of the information state assessments provide about student learning continues to motivate states to design alternative assessment exercises for use in their state assessment programs. Chart 3-2a shows the most commonly used types of nontraditional items or tasks in language arts and writing. Extended-response open-ended items are by far the favorite means of assessing writing, while language arts is assessed most often with enhanced multiple-choice items, short open-ended items, and extended response open-ended items. Chart 3-2b shows the most common exercise types for mathematics and science. Short open-ended exercises are used most commonly with mathematics

BEST COPY AVAILABLE



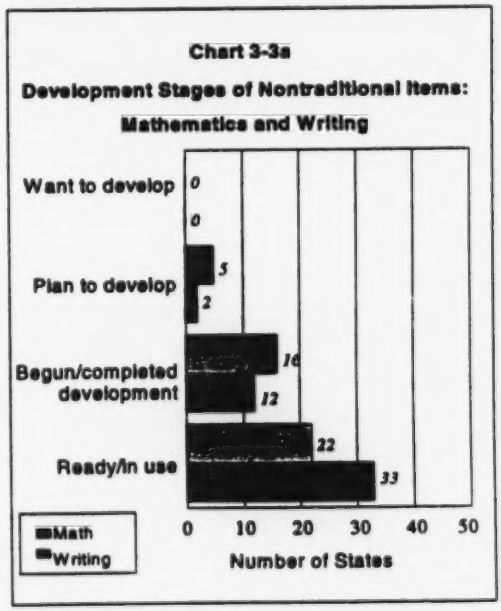


with extended response, individual performance assessment and enhanced multiple-choice exercises following. Science shows a similar pattern to mathematics.

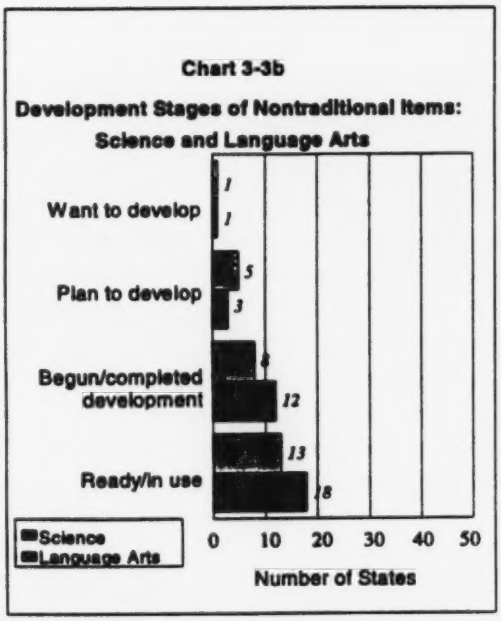
In 1992-1993, 40 states were creating or planned to create non-multiple-choice items in the five most commonly assessed subjects. In 1993-1994, 42 states are continuing with their nontraditional assessment activity; two others plan to take action in this area in the next three years. Thirty-four of these states report development efforts in subjects other than writing, while six states are working only in writing. The great variety of the work being done is accentuated by the fact that eight states report working on development of assessment items other than the 12 customary types listed in the survey.[6]

[6] The glossary at the back of this report defines each type of nontraditional assessment used in the survey and mentioned in this report.

Chart 3-3a shows how far along states are in the development of nontraditional items in writing and mathematics, the two subjects in which most developmental activity is occurring. The chart strongly suggests the states doing development work in these two subjects are, for the most part, well advanced. Most of the



BEST COPY AVAILABLE



items in these two subjects are ready for use or in use. There seem to be very few states beginning development work in these subjects.

In science and language arts (Chart 3-3b), the pattern suggests states are less far along. The differences between the number of states in the four stages is much smaller. There are relatively more states at the lower rungs of the development process. The principal difference between Charts 3-3a and 3-3b, however, is the smaller number of states with items ready for use or in use in science and language arts.

Constraints on Developing Nontraditional Assessment Items

Every form of assessment provides benefits and trade-offs. Traditional assessments are relatively inexpensive, easy to administer and score, and time-efficient. However, they have been criticized for focusing on what's easiest to assess—rote memory and isolated skills. On the other hand, while alternative forms of assessment provide students with the opportunity to demonstrate their ability to apply what they have learned, there are also trade-offs. Twenty-three of the 42 states which are developing non-traditional items reported that they encountered major difficulties. Twelve states reported that time was a major constraint, 15 indicated cost was the limiting factor and nine reported a lack of technical resources. Their responses pointed to the following issues, among others:

- Time. There are two time constraints. The first is the time to develop a test. This is compounded by a sense of urgency: several states report legislative mandates to put their programs into place before the tests were ready. The second constraint is the time to administer an alternative assessment in the classroom. In the time it would take a student to complete one or two performance assessments, that same student could have completed 200-items on a multiple-choice test.
- Cost. Again, there are several issues. Since the technologies are new, the procedures to develop items or tasks are not nearly as certain as in the development of NRTs. It takes more persons more time to develop and test such items. The time testing requires in the classroom adds to the cost of alternative assessment. Alternative assessment items are more expensive to score than multiple-choice tests. Alternative assessments require teachers or other professionals to record observational data or make judgments about extended artifacts of student performance. This requires the

skill and time of individuals if the work of many students is to be assessed. Professional development is also a very considerable expense for alternative assessment: staff need to understand the changes, staff need training in the consistent conduct and use of alternative assessment items, and staff need support in using and reporting the results of alternative assessment.

- Technical Quality. Because nontraditional items are a new technology, it is far from easy to obtain uniform results. While some technical concerns are not unique to nontraditional items and may in fact pose less of a threat, i.e., the issue of validity, they remain real, and others, such as reliability or generalizability, continue to be daunting. Traditional assessment often could move these concerns to the backroom and to the psychometric specialists for resolution. For alternative assessment, this is often not possible since the direct involvement of the teacher and student is much greater.

Writing Assessment

The most common form of nontraditional assessment has existed since the early 1970s, when the National Assessment of Educational Progress (NAEP) introduced writing assessment. Writing assessment is the most popular form of nontraditional assessment being used in state assessment programs. As pointed out in the previous section, and in Chart 3-3a, the developmental pattern for writing assessment is more advanced than for any other form of nontraditional assessment, although mathematics is catching up. This year, 38 states reported having a writing sample as part of their assessment program.

States most typically assess writing at three or four grade levels: grades 4, 5, or 6; grade 8; and grades 10 or 11. Most states test all students at each of the grade levels, although three states report testing only samples of students. Seven states report that the writing prompts are sampled within grade levels. The vast majority of states score one sample per student, with six states assessing two; two states assessing two or three; and only Kentucky, Massachusetts, and Vermont routinely assessing more than three samples as part of a portfolio. Most states require students to respond on demand, during a specified period of time (measured in minutes and hours), while 11 states allow an extended response period, measured in days and weeks. Most of the states that assess writing include all students eligible for assessment. Two states have a voluntary writing assessment program: Alaska's program is voluntary for students and Utah's is voluntary for schools or districts.

The two most common scoring methods used with writing samples are analytic scoring (providing scores on specific writing outcomes such as focus, organization, persuasiveness, and grammar) and holistic scoring (providing a single score based on the overall quality of the writing). Nineteen states report using holistic scoring, 5 use analytic scoring, and 11 use both. Twenty-three states report allowing students to revise their writing sample, but all who do so score only one (usually the final) revision. The number of states that allow revisions grows each year, indicating that the assessment of

writing is becoming more closely aligned with the way the writing process is taught.

Summary

Newer forms of nontraditional assessment are becoming increasingly popular in state assessment programs. Although implementation of nontraditional methods are complex, costly, and require new technology, almost all of the states are involved in some nontraditional assessment activity. Extensive research is needed and this, combined with the increased costs for this form of assessment, may stall full-scale implementation. It is clear from states' extensive experimentation with alternative forms of assessment that they are very interested in their potential benefits for educational reform purposes.

States know that "one size fits all" assessment does not exist. The use of any form of assessment involves trade-offs, and states are using a variety of assessment strategies to minimize the complications. States are expanding their assessment programs to include nontraditional assessment components to complement their existing traditional components. Massachusetts, Maryland, and Kentucky are the only states designing assessment programs that are exclusively nontraditional, although assessment programs in Arizona, California, and Vermont are predominantly nontraditional in focus. Still, most states are using both traditional and nontraditional assessment.

The purposes states assign to non-traditional assessments mirror those reported for traditional assessments: instructional improvement (32 states) and school performance reporting or accountability (30 states). The conflict which exists between these two purposes, as described in Chapter 2, also exists for these newer forms of assessment.

Much of the activity in the area of nontraditional assessment is still in writing, although mathematics is catching up. Activity in the other subject areas, most notably reading and language arts, science and social studies, is still in the developmental stages, and states appear to be moving cautiously toward implementation.

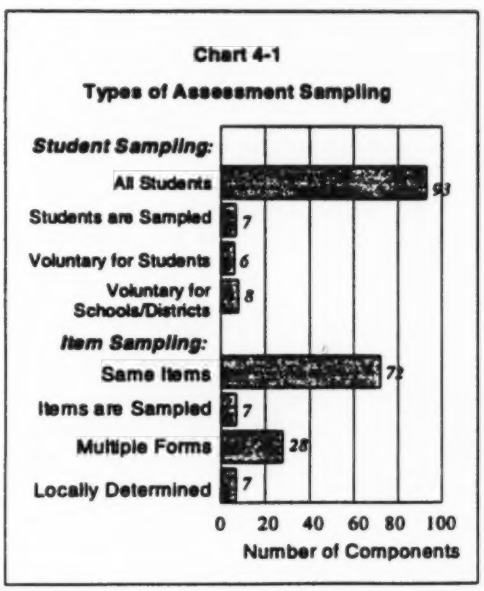
Chapter Four

Additional Assessment Issues

The annual survey included several questions concerning other important topical assessment issues. These questions included sampling issues, calculator use, policies regarding special populations, and the process states use to develop assessments.

Sampling

The survey asked the state assessment directors to report on the sampling of students and/or items in the state assessment program. Considerable variation was reported within, as well as among, states for two reasons: 1) many states have more than one assessment component and 2) several state assessments use different sampling techniques for different parts of an exam. Our analysis is based upon the 112 assessment components identified within the 45 states that conducted an assessment program in 1993-94. Chart 4-1 summarizes the sampling techniques states employ in their assessment program.



The most common practice, used in 42 states, is to assess all students at a given grade.[7] Ninety-three different testing components are conducted in this manner. In six states assessments are voluntary for schools or districts for eight components, and three states have components that are voluntary for students.

The most common sampling pattern for items is to give the same items to all students taking the test. This occurred in 71 of the assessment components, while multiple forms containing different items were used with 28 assessment components in 19 states.

The purposes of assessment influence a state's decisions regarding sampling strategies. Census sampling of students is often used if the exam is used to determine individual student proficiency. If, on the other hand, the assessment is used for school or program evaluation, data on individual students are not needed. In the seven states that use item sampling, however, group assessment (i.e., school or

[7] The grade is the usual target group. In some instances, particularly in high schools, the target group could be subject or course specific.

BEST COPY AVAILABLE

program evaluation) is usually the purpose. Different but equivalent forms of the exam may be used in testing situations that hold high stakes or consequences for individual students so that students cannot copy from one another. This also ensures that students who take a repeat or make-up exam will be presented with different items than those offered during the first administration.

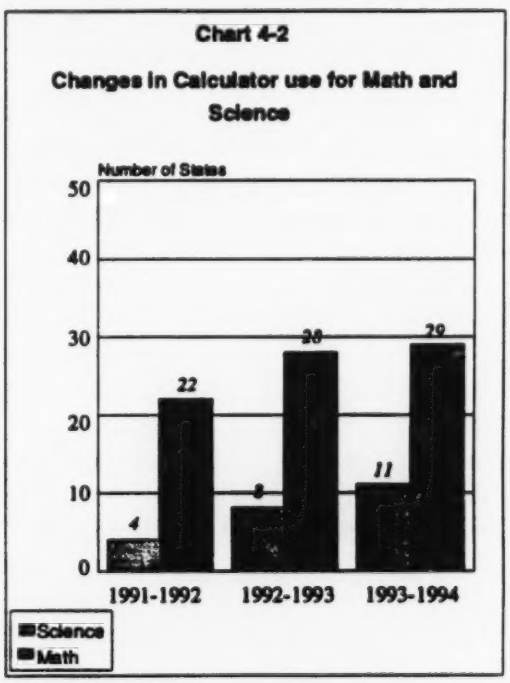
Variations of student and item sampling can be used together. For example, if a group score is desired, both students and items may be sampled. Student sampling calls for the assessment to be given to different but equivalent sets of students (e.g., a random sample of fourth graders in schools across the state). Within this group of students, the items are distributed over all of the students in the group so that each student receives only some of the assessment items. This form of sampling is used in eight states, usually in assessment situations with high stakes for the educators administering the exam. The randomness prevents teachers from teaching to the test, as they do not know which students will be given which items. Often, this combination of student and item sampling is used when the assessment results are needed for school or program accountability. These techniques may also be used when nontraditional forms of assessment (which have larger developmental and scoring costs) are used. Finally, if a group score (such as the overall score of a school or school district) is all that is needed, student or item sampling may be employed.

The most unusual form of state assessment allows school districts to develop or select their own assessment.[8] Six states follow this pattern to meet state reporting purposes.

Calculator Use

Chart 4-2 summarizes the use of calculators for statewide mathematics and science assessment. More than two-thirds of the states with mathematics assessments report the use of calculators, up two states from 1992-1993 and up seven from 1991-1992. Only seven of the states require calculator use, while the rest permit or encourage them.

Approximately half the states allow calculator use on all parts of the assessment, while the others allow its use on only a part. Fifteen states do not allow



[8] While now rare, this was an historically common pattern, predating state mandates for specific assessment practices. In Iowa, for instance, which reports no state assessment program, nearly all districts send annual reports of student testing to the SEA for compilation and analysis.

BEST COPY AVAILABLE

calculator use. With the National Council of Teachers of Mathematics recommending the use of calculators on "authentic" problems, this increase in the number of states allowing calculator use is encouraging.

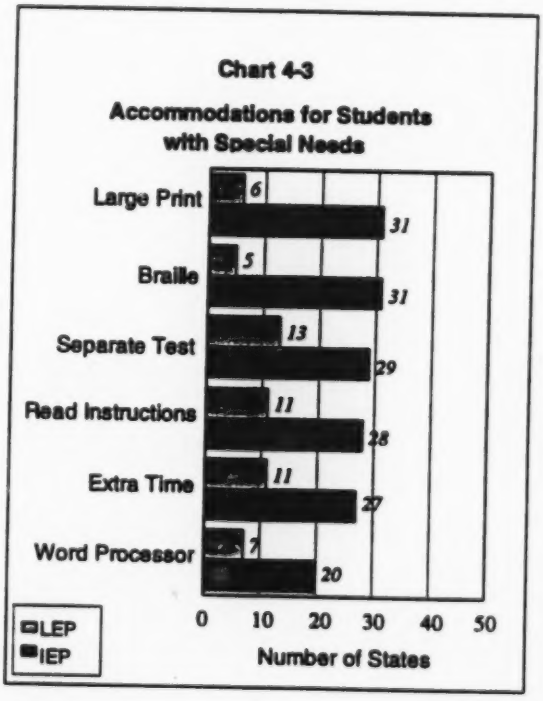
The use of calculators on state science assessments is much less frequent than for math, but their use with science assessments is also increasing sharply. In 1991-1992 four states reported calculator use in science; in 1992-1993 there were eight. This year 11 states report their use. No states require calculator use on their science assessments, but six permit their use, and five states encourage their use.

Despite this growth, almost one-third of the states have not yet embraced the use of calculators on their state assessments. This fact appears to contradict the trend toward alternative assessments to allow for complex, multiple-step problem solving. This may be due to the difficulty and expense of ensuring uniform use of calculators on assessments; i.e., that all students have the same opportunity to use the same kind of calculator.

Assessment of Special Populations

In most states, a special education student is included or excluded from the state assessment based on the recommendations of the Individualized Education Plan (IEP). More specifically, under federal special education law, parents have the right to determine whether or not their child will participate in the state assessment program. In a few states, the determining factor for inclusion is not the IEP but whether or not the student is reading at grade level. A number of states, including California, Idaho, Michigan, and Utah, use the 50 percent rule (if the student spends 50 percent or more of his or her time in regular education classes, the student is included in the state assessment), but even in these states the IEP may exclude the student. In Kentucky, students who cannot function in the regular curriculum may participate in an "alternative portfolio" assessment. High school proficiency or graduation tests also rely on the IEP, but a student who does not take or pass the state exam usually is denied a regular high school diploma.

Chart 4-3 summarizes the following findings: Accommodations for special education students appear to be more common than for Limited English Proficiency (LEP) students. In fact, while large print and Braille versions of the test were common for special education students with vision problems, only nine states allowed LEP students to take the test in their native language. In most cases



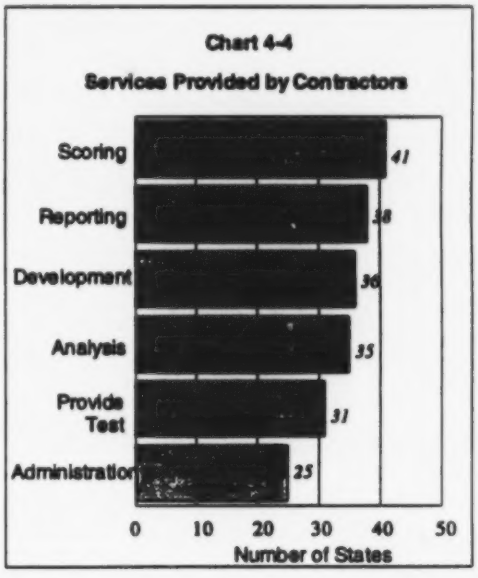
this was in subjects other than reading. Some states, such as Maryland and Hawaii, provide numerous accommodations, including reading and/or transcribing the test, extended time periods, small group administration, audiotaped versions, signed versions for the hearing impaired, use of calculators and/or word processors, large print, and Braille. A few states mentioned that decisions concerning special accommodations depended on the impact on validity (for example, students would not be read a reading test). In a number of cases, such as in Indiana and Virginia, the scores of students who receive special accommodations are flagged and excluded from the aggregate score for the district and/or school. Montana allows special education students' scores to be excluded from the district average in the area(s) in which special education services are provided.

Inclusion of LEP students in the state assessment program was treated differently from the inclusion of special education students, and more variety in approach existed across states. Two common approaches were noted. In most states, a determination is made about the English proficiency of the student. If the student is determined to be unable to read English, he or she is not required to take the test. The determination of English proficiency is made in a number of ways. The number of years in an English-as-a-second-language program is used in many states, including Florida, Idaho, and Massachusetts. Several states reported that the length of time that the student has been in the United States is also considered. A few states, such as Nevada, use a language test to determine language proficiency.

Developmental Process for State Tests

State assessment agencies received assistance, contracted and otherwise, from a variety of sources. The most common sources were the major test publishers, commercial scoring and reporting specialists, universities, and private consultants. The degree of involvement of these contractors varied considerably. A few states in effect depended entirely on a commercial contractor to design and conduct their assessment program. More typically, a commercial vendor handled the logistic of assessment, e.g., scoring and reporting, while the SEA did design and analysis work. Independent contractors, universities, and the commercial test publishers all assisted states with development work.

Chart 4-4 provides a tally of these involvements. Its most striking implication is the apparent widespread collaboration



BEST COPY AVAILABLE

across public and private, for-profit and commercial, enterprises in statewide assessment programs. There does, however, seem to be a growing shift towards the employment of smaller, independent, technically sophisticated consultancies, some private contractors, some housed in universities. One might assume that the experimentation with nontraditional forms of assessment, an area not well understood, may be the reason that so many states are seeking this assistance.

Chapter Five

History and Trends in Statewide Assessment

This is the third year in which the information collected on large-scale assessment programs at the state level has been collected systematically and made available. It is now possible to begin to see trends in the information. Although this report is based on three years of data, trends still must be interpreted cautiously since changes in student assessment programs take several years to conceptualize and implement. It is unlikely that substantial change will take place in the short run; however, the information reported here is similar to information collected less formally in the past, so that it is possible to combine current information with past information to perceive longer-term trends.

The purpose of the following sections is to comment on some of the changes that have occurred in the past 15 years. In addition, several issues that may imply future changes in assessment are mentioned.

Criterion-Referenced Assessment and Minimum Competency Tests

When the Association of State Assessment Programs was formed as an organization representing the assessment programs at the state and national level in 1977, two strong innovations had occurred and were being spread throughout the states. First, states such as Michigan had adopted a new form of measurement called "criterion-referenced tests." Rather than comparing student (or school or district) scores to national norms, scores were reported as pass-fail for individual objectives as well as a proportion of the outcomes passed. Second, tests were used to determine whether students had learned enough to receive a high school diploma. This use of minimum competency testing for high school graduation was exemplified by a landmark program in Florida. Early ASAP meetings were filled with discussions about the procedures for developing criterion-referenced tests, as well as surviving the inevitable legal challenges to the minimum competency tests, since the landmark legal case *Debra P. v. Turlington* was occurring at that time.

The predominant form of large-scale assessment at that time was norm-referenced tests. Interest in criterion-referenced tests was pushed along not only by the states that had adopted them as a form of assessment, but also by NAEP in its early years, since several states (such as California, Connecticut, Minnesota, and Wyoming) gave the early NAEP assessments in "piggyback" style in order to obtain state and national data on their students. Not only did this practice introduce these states to criterion-referenced testing, it also served as an introduction to the concept of the state NAEP assessment program.

Advent of Writing Assessment

In the 1970s, assessment was limited usually to mathematics and reading, with performance assessments just beginning in the area of writing. The NAEP assessments of writing in the early 1970s had encouraged the belief that having all students at one or more grade levels actually write essays would be feasible. Although more expensive than the much more prevalent multiple-choice tests of "writing," essay tests were thought to be more content valid, and it was believed that they would lead to better teaching of writing. However, strong debates about this concept occurred in the 1970s.

Expansion to Other Subject Areas

In the 1980s, new states adopted large-scale assessment programs as a tool for school reform and improvement. Each year at the ASAP meetings, one or two states new to large-scale assessment efforts would attend. In addition, states were beginning to add other subject areas to their assessments. They began to develop assessments in areas such as science, social studies (or one or more of its components, such as history or geography), health education, physical education, the arts, and vocational education. Interest also grew in sharing assessment items or tasks among the states. Attempts were made to create item banks among the states, but these generally proved to be unsuccessful since each state clung to its own set of student expectations, making sharing of corresponding items challenging at best.

Performance Assessment

The latter part of the 1980s also brought attention to performance assessment. Multiple-choice tests were (and still are) the major form of assessment used in most states, except for states that used a writing essay test. In the last few years, several trends have begun to occur. First, a small group of states (Maryland and Arizona being first, now joined by Maine, Massachusetts, and Delaware) developed and implemented entirely open-ended assessments of all students in several subjects at several grades. These states proved that it was feasible to administer alternative forms of assessment in a relatively cost-effective manner.

Second, some states are working on or piloting alternative forms of assessment. This work includes performance assessments given to individuals or small groups of students, examples of curriculum-embedded tasks in which assessment is intricately interwoven within teaching and is collected over several weeks or months, portfolios that collect examples of student work for later scoring, and other innovative forms of assessment.

As the survey indicates, few states have actually implemented these innovative alternative forms of assessment, but given the number of states reporting such work, it is logical to assume that these numbers will increase. It is likely that, given the costs of alternative assessment in money and time, most states will move toward the concept of an assessment system with different forms of assessment being used at different levels. For example, large-scale, standardized assessments with some alternative approaches might be used for state-level reporting, while more extensive programs of performance and/or portfolio

assessment might be used to meet school and/or classroom assessment needs. Hence, several states report that such innovative performance assessments are being developed for use by local educators.

Professional Development on Assessment

Attention to the forms of assessment used at both the state and local levels has encouraged another trend at the state level. As state-level educators have debated the form(s) of assessment appropriate for the state to use, increasing attention has been paid to the training of classroom teachers to collect and use information that might be gathered from innovative approaches to assessment within their classrooms. This trend is actually the convergence of several trends, including changes in student expectations to emphasize thinking and problem-solving skills (while de-emphasizing memorization of content knowledge), and support to alternative approaches to assessment such as projects, exhibitions, demonstrations, and the use of portfolios. The result is that many local districts and some state agencies are now providing classroom teachers with assessment learning experiences that teachers can apply in their classrooms. This attention to professional development on assessment for classroom teachers is particularly appropriate given that few if any teachers receive much in the way of preservice training on assessment.

Norm-Referenced Tests

When the ASAP group began meeting in 1977, the most commonly used assessments were commercially available (off-the-shelf) norm-referenced tests. Despite the attention to alternative forms of measurement, which is even more widespread today than it was 20 years ago, it is interesting to note that norm-referenced tests are still the predominant form of large-scale assessment in the United States. The trend in recent years has been a slight decrease in the use of norm-referenced tests at the state level. Several states that once emphasized such assessments have stopped doing so (in 1993, 31 states used norm-referenced tests, while 30 reported using these assessments in 1994).

There had been an expectation that this number would fall even further, given the de-emphasis on norm-referenced assessments in the Improving America's Schools Act (IASA), the reauthorization of Elementary and Secondary Education Act. States are no longer required to use such assessments for the evaluation of Chapter I compensatory education programs, nor for the monitoring of individual Chapter I student selection or evaluation. This was a major change in the legislation, which advocacy groups and others fought for and won. In place of such tests, states are required to develop and operate "comprehensive assessment systems" capable of reporting whether individual students and school programs are making "adequate yearly progress."

Two events conspired to confound this prediction. First, the November election brought to power at the state level chief state school officers, state board of education members, legislators, and governors with strongly held ideas about student standards and assessment. These ideas were oftentimes contrary to the spirit

BEST COPY AVAILABLE

of using new forms of assessment to raise standards. Given problems in some of the assessment efforts first implemented (in Arizona, California, Georgia, and Maine, to name a few), policymakers pushed to set aside innovative approaches to assessment and to return to commercially-available norm-referenced tests. While such debates and changes are too recent to be picked up even in the 1994 survey, they bear watching in the future.

Second, the changes implemented in the IASA legislation have proven to be less far-reaching than originally thought. Due to political changes in Washington, D.C., states will be required to change their statewide assessments substantially less than originally thought. States, for example, have five to six years to develop permanent comprehensive assessment systems (in only mathematics and reading, not in all of the national goal areas, unless they do so for all students). In the interim, transitional assessments of any type (norm-referenced, criterion-referenced, or performance assessments) can be used at state choice, so long as they are deemed to "measure challenging state content standards," which is left poorly defined in the federal legislation.

For these reasons, as well as because many policymakers desire to have comparative data on instruments developed outside the state, it is likely that norm-referenced tests will continue to be a major type of assessment being used in states. To satisfy this desire for normative information, but using measures of higher-level standards, some states (such as Kentucky and North Carolina) have administered the National Assessment of Educational Progress assessments to samples of students taking their statewide assessments in order to provide NAEP-like scores to buildings and districts (as well as the state). This recent innovation in providing normative information has the promise of allowing states to pursue new forms of assessment while still providing external referents for scores on the statewide assessments. It will be interesting to monitor the success of these efforts and to determine if this becomes a trend for the future.

National Efforts at Joint Development

Another trend is worth noting. Until 1990, most assessment development was carried out by individual states working alone or with the assistance of a contractor. Since then, two innovations in collaboration among the states have taken place. The first is the New Standards Project, co-directed by the University of Pittsburgh and the National Center for Education and the Economy, which has been working with a number of states and local districts to design and develop an innovative assessment system that will encourage thoughtful student learning in areas such as mathematics, language arts, and science. The second is the Council of Chief State School Officers' State Collaborative on Assessment and Student Standards (SCASS), which has nine projects in which states work together to develop innovative student assessments. Both of these activities mark a first for collaboration among the states. The states are actively working together to develop assessments from which they share and use the products, rather than simply exchanging information about innovative assessment approaches, as was the case in the past.

Future Issues and Their Impact on State Assessment

Overall, an examination of the changes in large-scale assessment programs during the past 20 years shows a substantial change in the number of states with such programs, the subject areas assessed, and the types of assessment measures used—as well as the types of assessment measures being developed (and the manner in which this development is proceeding). These changes have only increased in the past few years with the considerable public attention paid to the quality of schools. Not surprisingly, these changes have led a number of states to reexamine assessment program designs that were adopted in years past. A number of states are examining whether their current assessment designs are still adequate and are looking at how such recent programs such as NAEP, the New Standards Project, and SCASS fit within their overall assessment design. Given the number of states that are conducting such examinations, further changes in the nation's large-scale assessment programs are likely.

Several trends appear at the state and local levels that may have a long-term impact on the shape of large-scale assessment programs at the state level. Certainly, the current emphasis on performance or alternative assessments is not going to disappear. Although there have been some successes (such as in Maryland and Kentucky), the setbacks in California, Arizona, Indiana, and elsewhere indicate that widespread acceptance of performance assessment is certainly not automatic. Technical issues need to be addressed in a sound manner, and policymakers and the public need to understand the reasons for such measures, the student expectations they measure, and the reasons why both traditional and performance assessments are needed. States and others interested in innovative forms of assessment will need to make sure important parties are "on board" before engaging in this innovative development work.

Certainly, there will be some impact from the drive now under way in some states to "deregulate" public education and return control of it to local school districts. While this drive is taking several forms, it would not be unexpected for these pressures to affect the extent and types of student assessment in the future. In some states, this may mean less attention to statewide student expectations and measures, while in other places, it may mean just the opposite.

The pressure to provide appropriate assessment training and experiences to classroom teachers is also not likely to abate. The collaborative work across states is likely to spread innovative approaches to assessment more quickly than it has in the past. In addition, the outside political pressures to use assessment as a tool for reform of schools is not likely to lessen. Changes brought about by federal legislation such as Goals 2000 and IASA will occur as well, but perhaps at a slower pace than once thought. In addition, it is uncertain how the battles between chief state school officers and governors shaping up over control of education funds in federal block grant programs will affect large-scale student assessment programs.

Finally, the reauthorization of the NAEP program has brought several changes that also may affect states. In recent years,

BEST COPY AVAILABLE

NAEP has offered the trial state NAEP programs, but, unfortunately, recent appropriations for the program have not permitted a full-scale state NAEP program to be offered. If the program is funded at a higher level, it might affect the number of states that administer norm-referenced tests to students at one or more grade levels, since the NAEP data provide the types of national comparisons that states desire that are more current, less expensive, and more technically sound than many traditional norm-referenced tests.

Many swirling, cross-cutting trends at the state level are affecting large-scale assessment programs, and it is likely that these trends will occur in the future. With the State Student Assessment Program database, it should be easier to track the course of changes in large-scale assessment programs at the state level. Future editions of this report will begin to indicate more precisely just how such changes are occurring.

APPENDICES

Definition of Terms

Computer Adaptive Testing is any assessment, other than multiple-choice questions or worksheets, that requires the student to respond to the assessment items or task with the aid of a computer. For example, the student responds to several questions to determine his or her ability and then is moved into the performance task that best meets the student's ability level.

Enhanced multiple-choice is any multiple-choice question that requires more than the selection of one correct response. Most often, the task requires the students to explain their responses.

Extended-response, open-ended indicates any item or task that requires the student to produce an extended written response to an item or task that does not have one right answer (e.g., an essay or laboratory report).

Group performance assessment is any assessment that requires students to perform the assessment task in a group setting. For example, a performance assessment as defined in individual performance assessment becomes a group performance assessment when the task is performed in a group and the individual's rating is based on his or her performance as part of the group.

Individual performance assessment is any assessment that requires the student to perform (in a way that can be observed) an assessment task alone. For example, a student may be asked to perform a laboratory experiment or carry out a community service project and write about the results. The performance of the laboratory experiment and the community service project makes this assessment an individual performance assessment versus an extended-response assessment, when the quality of the performance itself and not just the quality of the writing is rated.

Interview is an assessment technique in which the student responds to verbal questions from the assessor.

Nontraditional test items indicate any assessment activity other than a multiple-choice item from which the student selects one response. These items or performances are rated using an agreed-upon set of performance criteria in the form of a scoring guide or a scoring rubric or in comparison to benchmark papers or performances.

Observation is an assessment technique that requires the student to perform a task while being observed and rated using an agreed-upon set of scoring criteria.

Opportunity to learn refers to the educational approaches that are necessary to provide students with the "opportunity to learn" the standards on which they are being assessed; unlike student standards, "opportunity to learn" standards hold the school accountable for providing these learning opportunities to students.

Portfolio is an accumulation of a student's work over time that demonstrates growth toward the mastery of specific performance criteria against which the tasks included in the portfolio can be judged.

Project, exhibition, or demonstration is the accomplishment of a complex task over time that requires demonstrating mastery of a variety of desired outcomes, each with its own performance criteria, that can be assessed within the one project, exhibition, or demonstration.

Short-answer, open-ended is any item or task that requires the production of a short written response on the part of the respondent. Most often, there is a single right answer (for example, a fill-in-the blank or short written response to a question).

BEST COPY AVAILABLE

Summary Table

This table summarizes a significant amount of information from the SSAP database and is somewhat complex. Please keep the following in mind when reading the table.

Most states conduct several assessment programs side-by-side (labeled #COM, for components). This table aggregates across these components. It should be read, emphasizing the term "at least" in the following sense: Alaska conducts at least one program assessing all fourth or sixth or eighth graders in language arts or math or writing; it also assesses at least some fifth and tenth graders; Alaska makes use of a norm-referenced multiple-choice test and a writing sample; these assessments are conducted to diagnose or place students, to improve instruction, to evaluate programs, and to generate reports on school performance.

This table is distilled from the 75 SSAP tables. The single exception is North Carolina. The summary table lists four components for the state. The tables in Part 3 of the data base list only one component. The description for the single component is correct; however, there are three other components in the North Carolina program. This corrected information is reflected in the Summary Table.

SUMMARY TABLE: STATE ASSESSMENT PROGRAMS, AS OF OCTOBER 1994

ST	#COM	Grades K	1	2	3	4	5	6	7	8	9	10	11	12	Subjects L	M	W	SS	Sc	Voc	Ap	Rd	Test Type NRT	CRT	Per	Prt	WS	Assessment Purpose Kgl	StD	Imp	PE	StP	Gr	ED	HD	StA	SPR	Acr	ScA	SK	School Consequences FG	FL	AL	ER	PWL	W	T	D	Oth
AK	2	-	-	-	-	A	V	A	-	A	-	V	-	-	A	A	V	-	-	-	-	-	A	-	-	-	V	-	Y	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
AL	7	-	-	A	A	A	A	A	A	A	-	A	A	-	A	A	A	A	A	-	-	-	-	A	-	-	A	-	Y	Y	Y	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
AR	2	-	-	-	A	A	S	-	A	A	S	A	-	-	A	A	A	A	A	-	-	-	-	A	-	-	A	-	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	-	-	-	-
AZ	2	-	-	-	A	A	-	-	A	A	-	A	A	A	A	A	A	-	-	-	-	-	-	A	-	-	A	-	Y	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
CA	3	-	-	-	-	A	A	-	-	A	V	A	V	V	A	A	A	A	A	V	-	-	-	A	-	-	A	-	-	Y	Y	-	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-
CO	0*	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	A	-	-	-	-	-	-	V	V	A	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
CT	2	-	-	-	-	A	-	A	-	A	-	A	-	-	A	A	A	-	A	-	-	-	-	-	A	A	-	A	-	Y	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
DE	2	-	-	-	A	-	A	-	-	A	-	A	-	-	A	A	A	-	-	-	-	-	-	A	-	A	-	A	-	Y	Y	-	-	-	-	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-
FL	3	-	-	-	A	A	-	-	-	A	-	A	A	A	A	A	A	-	-	-	-	-	-	A	-	-	A	A	-	Y	Y	-	-	-	-	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-
GA	6	A	-	-	A	-	A	-	-	A	-	A	A	A	A	A	A	A	A	-	-	-	A	A	-	-	A	A	-	Y	Y	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
HI	3	-	-	-	A	-	-	A	-	A	V	A	V	V	A	A	-	-	-	-	-	-	-	A	V	-	-	-	-	Y	Y	-	Y	-	-	-	Y	-	Y	-	-	-	-	Y	-	-	-	-	-
IA	0	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	-	-	-	-	A	-	-	-	-	-	-	Y	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-
ID	2	-	-	-	-	-	-	A	-	A	-	A	-	-	A	A	A	-	-	-	-	-	-	A	-	A	-	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
IL	1	-	-	-	A	A	-	A	A	A	-	A	-	-	A	A	A	A	A	-	-	-	-	A	-	-	-	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
IN	1	-	-	A	A	-	-	A	-	A	A	-	-	-	A	A	-	A	A	-	-	-	-	A	-	-	-	A	-	-	Y	-	Y	-	-	-	Y	-	-	-	-	-	-	Y	-	Y	-	-	-
KS	1	-	-	-	A	A	A	-	A	A	A	A	-	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	Y	-	-	-	Y	-	-	-	-	-	-	Y	-	Y	-	-	-
KY	3	-	-	-	A	A	A	-	-	A	-	-	A	A	A	A	A	A	A	A	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
LA	4	A	-	-	A	A	A	A	A	-	-	-	A	-	A	A	A	A	A	-	-	A	-	A	-	-	A	A	-	Y	Y	-	Y	-	-	-	Y	-	-	-	-	Y	-	-	Y	Y	-	-	-
MA	0*	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	A	A	A	-	-	A	-	A	-	-	A	A	-	Y	Y	-	Y	-	-	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-
MD	3	-	-	-	A	-	A	-	A	A	A	-	-	-	A	A	A	A	A	-	-	-	-	S	A	A	-	A	-	-	Y	-	Y	-	-	-	Y	Y	Y	Y	Y	-	Y	-	Y	-	-	-	-
ME	1	-	-	-	-	A	-	-	-	A	-	A	-	A	A	A	A	A	A	-	-	-	-	-	A	-	-	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
MI	2	-	-	-	-	A	A	-	A	A	-	A	A	A	A	A	-	-	A	-	-	-	-	-	A	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
MN	1	-	-	-	-	S	S	S	-	S	S	-	S	-	S	S	S	S	S	-	-	-	-	-	S	-	S	-	-	-	Y	-	Y	-	Y	-	Y	Y	Y	Y	-	Y	Y	-	-	Y	Y	-	-
MO	2	-	-	S	S	S	S	S	S	S	S	S	S	-	S	S	S	S	S	-	-	-	-	-	S	-	S	-	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
MS	3	-	-	-	A	A	A	A	A	A	A	-	A	-	A	A	A	-	-	-	-	-	-	A	-	-	A	A	-	-	Y	-	Y	-	-	-	Y	-	Y	-	-	-	-	-	-	-	-	-	-
MT	1	-	-	-	A	-	-	-	-	A	-	-	A	-	A	A	-	A	A	-	-	-	-	A	-	-	-	A	-	-	Y	-	-	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-
NC	4	-	A	A	A	A	A	A	A	A	A	A	A	-	A	A	-	A	A	A	-	-	-	A	-	-	A	A	-	-	Y	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-
ND	1	-	-	-	A	-	-	A	-	A	-	-	A	-	A	A	-	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	Y	-	-	-	Y	Y	-	-	-	-	-	-	-	-	-	-	-
NE	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NH	1	-	-	-	A	-	-	-	-	-	-	-	-	-	A	A	-	-	-	-	-	-	-	-	A	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NJ	2	-	-	-	-	-	-	-	-	A	-	-	A	-	A	A	A	-	-	-	-	-	-	-	A	A	-	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
NM	4	-	A	A	A	A	A	A	-	A	-	A	-	-	A	A	A	A	A	-	-	-	A	-	A	A	A	A	-	-	Y	-	Y	-	-	-	Y	Y	-	Y	-	-	-	Y	-	Y	-	-	-
NV	4	-	-	-	A	A	A	A	-	A	-	A	A	A	A	A	A	A	A	-	-	-	A	A	-	-	A	A	Y	Y	Y	-	Y	-	-	-	Y	-	-	-	Y	-	-	-	-	-	-	-	-
NY	7	-	-	-	A	A	A	A	-	A	A	A	A	V	A	A	A	A	A	V	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
OH	4	-	-	-	-	A	-	A	-	A	A	-	-	A	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	Y	Y	Y	Y	Y	-	-	-	-	-	-	Y	-	-	-	-	Y
OK	2	-	-	-	A	-	A	-	A	A	-	A	-	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	Y	-	Y	-	-	-
OR	2	-	-	-	A	-	A	-	-	A	-	A	-	-	A	A	A	-	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	Y	-	Y	-	-	-
PA	2	-	-	-	-	-	A	A	-	A	A	-	A	-	A	A	A	-	-	-	-	-	-	A	-	-	A	V	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
RI	3	-	-	-	-	A	-	-	-	A	-	A	-	-	A	A	A	-	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
SC	2	-	-	-	A	A	A	A	A	A	A	A	A	-	A	A	A	-	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
SD	2	-	-	-	A	A	A	-	A	A	A	A	A	-	A	A	-	A	A	A	A	-	-	A	-	-	A	A	Y	Y	Y	-	Y	-	-	-	Y	-	-	-	Y	Y	-	-	Y	Y	-	-	-
TN	4	-	-	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	Y	-	-	Y	Y	Y	-	-
TX	1	-	-	-	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	Y	Y	-	-	-	Y	-	Y	-	-	-	-	-	Y	Y	Y	-	-
UT	3	-	V	V	V	V	A	V	V	A	V	V	A	V	A	A	V	A	A	-	-	-	-	A	V	-	V	A	-	-	Y	-	-	-	-	-	Y	Y	-	Y	Y	-	Y	Y	Y	Y	Y	-	-
VA	2	-	-	-	-	A	A	-	-	A	-	-	A	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	-
VT	2	-	-	-	-	A	A	-	-	A	-	-	A	-	A	A	-	A	A	-	-	-	-	A	-	-	A	V	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	Y
WA	1	-	-	-	-	A	-	-	-	A	-	-	A	-	A	A	-	A	A	-	-	-	-	A	-	-	A	V	-	-	Y	-	-	-	-	-	Y	-	-	-	-	Y	-	-	-	-	-	-	-
WI	2	-	-	-	A	-	-	-	-	A	-	A	-	-	A	A	A	A	A	-	-	-	-	A	-	-	A	A	-	-	Y	-	-	-	-	-	Y	-	-	-	-	-	-	-	-	-	-	-	Y
WV	3	-	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	-	-	-	-	A	-	A	-	A	-	-	Y	-	-	-	-	-	Y	Y	-	Y	-	-	-	-	-	-	-	-	Y
WY	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TOTAL		2	4	8	25	31	25	23	17	42	19	21	32	7	44	45	37	27	30	5	2	3	32	31	25	7	38	4	26	43	37	6	17	3	3	10	41	12	7	10	8	5	11	5	13	9	7	2	4

Note: CO and MA suspended their assessment programs in 1993-94.

- CELL CONTENTS ---
A=All are tested
S=Sample is tested
V=Voluntary for students, schools, or districts
Y=Purpose applies to at least one component

---- SUBJECTS --------
L =Language arts, including reading
M =Math
W =Writing
SS =Social studies
Sc =Science
Voc=Vocational
Ap =Aptitude
Rd =Readiness

--- TEST TYPE ------
NRT=Norm-referenced test
CRT=Criterion-referenced test
Per=Performance testing
Prt=Portfolio assessment
WS =Writing samples

---------- ASSESSMENT PURPOSE ----------
Kgl=Kg or Grade 1 readiness
StD=Student diagnosis, placement
Imp=Improvement of instruction
PE =Program evaluation
StP=Student promotion
Gr =HS graduation
ED =Endorsed diploma
HD =Honors diploma
StA=Students awards, recognition
SPR=School performance reporting
Acr=School accreditation
ScA=School awards, recognition
SK =HS skills guarantee

--- SCHOOL CONSEQUENCES -
FG =Funding gain
FL =Funding loss
AL =Accreditation loss
ER =Exemption from regulations
PWL=Probation, watch lists
W =Warnings
T =Takeover
D =Dissolution
Oth=Other

#COM=Number of components

BEST COPY AVAILABLE

State Student Assessment Programs Database

ORDER FORM

I would like to order the following from the State Student Assessment Programs (SSAP) Database. Payment in the appropriate amount is enclosed, made payable to "NCREL." Indicate the quantities of each type of material desired:

___	A book listing the 1994-95 data tables	$ 29.95
___	The book on diskette (select PC type)	
	___ Macintosh	$ 14.95
	___ Windows,	$ 14.95
___	An Annual Report	$ 9.95
___	The 1995 Annual Report on diskette	
	___ Macintosh	$ 5.95
	___ Windows	$ 5.95
___	Computer Data Files in the format select	$100.00

$_____ Total Payment Enclosed (above prices include shipping costs)

Shipping Information:

Please fill out the shipping information completely to avoid delays in shipping the desired materials.

Name _____

Address _____

_____

Telephone _____FAX_____

If you need copies of previous updates or listings, call Dina Czocher, 708/218-1274.
If you need copies of the databases as computer files, call Arie van der Ploeg, 708/218-1076.

Return this order form, with payment to: NCREL, Dina Czocher, 1900 Spring Road, Suite 300, Oak Brook, IL 60521, phone number 708/218-1274 or FAX 708/218-4989 or CCSSO, Debra Roeber, 1664 Algoma Drive, Okemos, MI 48864, phone and fax number 517/347-1145

Council of Chief State School Officers
One Massachusetts Ave., NW, Suite 700 • Washington, DC 20001-1431 • (202) 408-5505 • Fax (202) 408-8072

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300 • Oak Brook, IL 60521-1480 • (708) 571-4700 • Fax (708) 571-4716

BEST COPY AVAILABLE